Finitely Additive Priors, Coherence, and

The Marginalization Paradox

by

William D. Sudderth

Technical Report No. 358

*Research supported by NSF Grant MCS77-28424.

ABSTRACT

The marginalization paradox, which arises in a Bayesian theory of inference based on improper, countably additive prior distributions, cannot occur when proper, finitely additive priors are used instead. Nor is there a paradox if Bayesian theory is based directly on a notion of coherence.

1. Introduction

The use of finitely additive probability distributions is, in some respects, similar to the use of improper, countably additive ones. Both types of distributions depart from the widely accepted Kolmogorov axioms and both types make possible within a Bayesian framework a number of standard statistical inferences which are impossible for a Bayesian restricted to the Kolmogorov framework.

As a simple example, consider a normal variable X with variance 1 and unknown mean θ. After observing $X = x$, a Bayesian with an improper, uniform prior for θ on the whole real line, has as posterior for θ a normal distribution with mean x and variance 1. A Bayesian whose prior for θ is a finitely additive, translation invariant measure on the line has the same posterior (Example 4.1 of [2]), but no proper, countably additive prior could result in it.

In addition to the similarities, there are also crucial differences between the finitely additive and improper theories. Points in favor of improper priors are the relative ease of computation and the essential uniqueness of the corresponding posteriors. On the other hand, improper priors can result in incoherent inferences whereas proper, finitely additive priors cannot do so (cf. Corollary 1 and Example 5.2 of [2]). Similarly, improper priors can lead to marginalization paradoxes (cf.[1] and [4]), but, as is the main content of this paper, no such paradox can arise in the finitely additive setting.

After a formulation of the paradox is presented in section 2, two demonstrations of its impossibility in the finitely additive theory are presented in sections 3 and 4, respectively. The argument of section 3 is indirect and relies on the connection between finitely additive priors and coherence established in [2]. This argument will actually show the marginalization phenomenon to be impossible in a theory based on coherence even if the definition of "coherence" is not precisely that used in [2]. The argument given in section 4 is a direct calculation which shows the paradox cannot arise in any theory based on proper distributions whether the distributions are required to be countably additive or only finitely additive.

2. The marginalization paradox

Let Θ and X be nonempty sets to be regarded as the parameter space and observation space, respectively. Let p be a conditional probability distribution on X given Θ; that is, p associates with each $\theta \varepsilon \Theta$ a probability distribution p_θ on a fixed σ - field of subsets of X. (Henceforth, to avoid verbosity, σ - fields will not be mentioned nor will assumptions of measurability be made explicit. However, all sets which occur are hereby assumed to be equipped with suitable σ - fields and all functions to be appropriately measurable.)

The scenario of the paradox revolves around two Bayesians B_1 and B_2. The first Bayesian, B_1, has a prior distribution π on Θ and is able to calculate his posterior q. (A posterior q is a conditional probability distribution on Θ given X, and q_x reflects B_1's posterior opinion after the observation x. In the improper theory, the density for q is calculated formally from Bayes formula, but, in the proper theory, q is calculated using the theory of conditional distributions where Bayes formula is not always available. A complete definition is in [2] or section 4 below.) It is assumed that B_1 is interested only in the parameter $\tau = \phi(\theta)$ which takes values in the set T. Thus, after an observation x, B_1 calculates q_x' which is the q_x - distribution of τ. It is further assumed that q_x' depends only on a statistic $y = r(x)$ which takes values in the set Y. That is, $q_x' = q_{x_1}'$ whenever $r(x) = r(x_1)$ Finally, it is assumed that p_θ', the p_θ - distribution of y, depends only on $\tau = \phi(\theta)$. To sum up, by a process of marginalization, B_1 is able to describe his inference about τ by a conditional probability distribution $\bar{q}$ on T

given Y where

$$\bar{q}_y = q'_x \tag{2.1}$$

whenever $y = r(x)$, and there is also a natural conditional distribution $\bar{p}$ on Y given T defined by

$$\bar{p}_\tau = p'_\theta \tag{2.2}$$

whenever $\tau = \phi(\theta)$.

Enter B_2. The second Bayesian is informed about B_1's inference represented by $\bar{q}$ and is able to calculate $\bar{p}$ from the given p. It seems plausible that $\bar{q}$ should represent a Bayesian inference relative to the reduced parameter space T and the reduced observation space Y. Thus B_2 should be able to find a prior $\bar{\pi}$ on T for which $\bar{q}$ is the posterior. If no such $\bar{\pi}$ exists, then the situation is called a marginalization paradox.

Here is an example due to Stone and Dawid [4],

<u>Example 1</u>: Suppose $\theta = (\mu,\sigma^2)$ and p_θ is the distribution of a random sample $x = (x_1,\cdots,x_n)$ from a $N(\mu,\sigma^2)$. Assume that the parameter of interest is $\tau = \mu/\sigma$. If B_1 has the improper prior $d\mu d\sigma/\sigma$, then, as shown in [4], his marginal posterior density for τ given x is proportional to

$$\exp\left(\frac{-n\tau^2}{2}\right) \int_0^\infty \omega^{n-2} \exp\left(\frac{-\omega^2}{2} + y\tau\omega\right)d\omega, \tag{2.3}$$

where $y = (\Sigma x_i)/(\Sigma x_i^2)^{\frac{1}{2}}$. It is also shown in [4] that the density for y given θ depends only on τ and is proportional to

(2.4) $$\exp\left(\frac{-n\tau^2}{2}\right)\left(1 - \frac{y^2}{n}\right)^{(n-3)/2} \int_0^\infty \omega^{n-1} \exp\left(\frac{-\omega^2}{2} + y\tau\omega\right)d\omega,$$

Since (2.4), considered as a function of θ, does not divide (2.3), there is no improper, prior density that B_2 can choose for τ which will enable him to reach the same posterior for τ as B_1.

Stone [3] points out that it follows from results of Heath and Sudderth [2] that B_1 could choose a proper, finitely additive prior for θ which would lead to the same posterior. However, as the next two sections show, B_2 can find a proper, finitely additive prior which leads to the same posterior for τ. In fact, by the Propostiion of section 4, if B_1 has the finitely additive prior π for θ, B_2 can take his prior for τ to be the π - distribution of τ.

3. Coherence

Consider a bookie B_1 who, conditional on the observation x, posts odds on subsets of Θ in accordance with the probability distribution q_x. Imagine also a gambler who, conditional on x, stakes $s(x) = b(x)\, q_x(A_x)$ on the subset A_x of Θ thereby winning $b(x)$ if A_x contains the true parameter value and losing $s(x)$ otherwise. If the function b is bounded, the pair $S = (A,b)$ is called a <u>simple betting system</u> and its expected payoff to the gambler as a function of θ is

$$E(\theta) = \int b(x)[A_x(\theta) - q_x(A_x)]dp_\theta(x),$$

where $A_x(\theta)$ is 1 or 0 according to whether θ is in A_x or A_x^c. The bookie B_1 or his odds system q is said to be <u>coherent</u> if there does not exist a finite number $S_1,\cdots,S_n$ of simple betting systems such that the sum $E_1(\theta) +\cdots+ E_n(\theta)$ of the corresponding payoff functions is uniformly positive as a function of θ. This definition of coherence is the one used in [2] where it is shown that B_1 is coherent if and only if q is the posterior for some finitely additive prior π on Θ.

Suppose now that B_1 is a coherent bookie so that q is the posterior of the finitely additive prior π. Assume also that B_1 decides to post odds only on those subsets of Θ which depend only on $\tau = \phi(\theta)$, which is equivalent to his posting odds on subsets of T in accordance with q_x'. suppose further that, as in section 2, q_x' depends only on $y = r(x)$ so that B_1 is effectively posting odds according to the $\bar{q}_y$ of (2.1) conditional on y. Finally assume that p'_θ depends only on τ so that $\bar{p}_\tau$ can be defined as in (2.2).

If a second bookie B_2 now decides to post odds on subsets of T according to $\bar{q}$, he will be coherent also. Intuitively, this is clear because B_2 is posting the same odds as B_1. To be more precise, let $S' = (A',b')$ be a simple betting system to be used against B_2. So A'_y is a subset of T for each y and b' is a bounded function from Y to the reals. Define the system $S = (A,b)$ by setting

$$A_x = \{\theta: \phi(\theta) \in A'_{r(x)}\}$$

and

$$b(x) = b'(r(x))$$

for every x. It is straightforward to verify that the corresponding expected payoffs satisfy

$$E(\theta) = E'(\phi(\theta)).$$

Thus, any finite collection $S'_1,\cdots,S'_n$ with a uniformly positive expected total payoff would result in a collection $S_1,\cdots,S_n$ with the same property. But B_1 is coherent by assumption and no such collection exits. Thus B_2 is coherent and, by the result of [2] mentioned above, $\bar{q}$ is the posterior for some finitely additive prior π.

Except for the connection to finite additivity, the argument just presented applies when the definition of coherence is modified in various ways. For example, suppose B_1 is said to be coherent if there do not exist $S_1, \cdots, S_n$ with $E_1(\theta) + \cdots + E_n(\theta)$ everywhere nonnegative and somewhere positive. Then B_2 is coherent in the same sense.

4. Finitely additive priors

The object in this section is to demonstrate that, under the assumptions of section 2 and the additional assumption that B_1's prior π on Θ is proper, then B_2 can use as prior on T the π - distribution of $\tau = \phi(\theta)$ and thereby arrive at the posterior $\bar{q}$ which corresponds to the distribution of τ under B_1's posterior q. For the demonstration precise definitions of the terms "marginal" and "posterior" are needed.

Let π be a proper, (possibly only) finitely additive probability measure on Θ, and let p be a conditional probability distribution on X given Θ. The marginal m of (π,p) is the probability measure defined on subsets A of X by

$$m(A) = \int p(A|\theta)\, d\pi(\theta), \tag{4.1}$$

where $p(A|\theta) = p_\theta(A)$. A conditional probability distribution q on Θ given X is a posterior for (π,p) if

$$\int\int h(\theta,x)\, dq_x(\theta)\, dm(x) = \int\int h(\theta,x)\, dp_\theta(x)\, d\pi(\theta) \tag{4.2}$$

for bounded, real-valued functions h defined on $\Theta \times X$. Thus q is a posterior if the pair (m,q) determines the same distribution on the product space $\Theta \times X$ as does (π,p).

Assume now the setting of section 2 and let $\bar{\pi}$ be the π - distribution of τ. That is, $\bar{\pi}$ is the probability measure on T which satisfies

$$\bar{\pi}(B) = \bar{\pi}(\phi^{-1}(B)),$$

for subsets B of T or, equivalently,

$$\int g(\tau)\, d\bar{\pi}(\tau) = \int g(\phi(\theta))\, d\pi(\theta)$$

for bounded, real-valued g defined on T.

Proposition: $\bar{q}$ is a posterior for $(\bar{\pi},\bar{p})$.

Proof: First calculate the marginal $\bar{m}$ of $(\bar{\pi},\bar{p})$ on Y as follows: for $A \subset Y$,

$$\begin{aligned}\bar{m}(A) &= \int \bar{p}(A|\tau)\, d\bar{\pi}(\tau) \\ &= \int \bar{p}(A|\phi(\theta))\, d\pi(\theta) \\ &= \int p'(A|\theta)\, d\pi(\theta) \\ &= \int p(r^{-1}(A)\,|\,\theta)\, d\pi(\theta) \\ &= m(r^{-1}(A)).\end{aligned}$$

so $\bar{m}$ is the m - distribution of $y = r(x)$ and satisfies

$$\int g(y)\, d\bar{m}(y) = \int g(r(x))\, dm(x) \tag{4.3}$$

bounded, real-valued g defined on Y.

Now let h be a bounded, real-valued function defined on $T \times Y$ and calculate again.

$$\begin{aligned}\iint h(\tau,y)\, d\bar{q}_y(\tau)\, d\bar{m}(\tau) &= \iint h(\tau,r(x))\, d\bar{q}_{r(x)}(\tau)\, dm(x) \\ &= \iint h(\tau,r(x))\, dq'_x(\tau)\, dm(x) \\ &= \iint h(\phi(\theta), r(x))\, dq_x(\theta)\, dm(x) \\ &= \iint h(\phi(\theta),r(x))\, dp_\theta(x)\, d\pi(\theta) \\ &= \iint h(\phi(\theta), y)\, dp'_\theta(y)\, d\pi(\theta) \\ &= \iint h(\phi(\theta), y)\, d\bar{p}_{\phi(\theta)}(y)\, d\pi(\theta) \\ &= \iint h(\tau,y)\, d\bar{p}_\tau(y)\, d\bar{\pi}(\tau).\end{aligned}$$

The first equality is by (4.3) and the fourth is by (4.2). The others are immediate consequences of the definitions.

The proof of the proposition is now complete.

References

[1] Dawid, A. P., Stone, M. and Zidek, J. V. (1973). Marginalization Paradoxes in Bayesian and Structural Inference. Journal of the Royal Statistical Society, Series B, 35, 189-233.

[2] Heath, David and Sudderth, William (1978). On finitely additive priors, coherence, and extended admissibility. Annals of Statist. 6, 333-345.

[3] Stone, Mervyn (1979). Review and analysis of some inconsistencies related to improper priors and finite additivity. To appear in the Proceedings of the 6th International Congress on Logic, Methodology and Philosophy of Science.

[4] Stone, M. and Dawid, A. P. (1972). Un-Bayesian implications of improper Bayes inference in routine statistical problems. Biometrika 59, 369-375.